
Contents

Introduction	XI
Basic Notations	XXIII
<hr/> Part I: Randomized Algorithms <hr/>	
1 Feedback, Averaging and Randomization in Control and Data Mining	3
1.1 Feedback	3
1.1.1 Information and Control	3
1.1.2 Signals and Data Sizes	6
1.1.3 Observations with Noise	6
1.2 Averaging	6
1.2.1 Data Averaging	6
1.2.2 Averaging in Stochastic Control	6
1.3 Efficiency of Closed-Loop Strategies under Uncertainty	6
1.4 Estimation under Arbitrary Noise	11
1.4.1 Can Smart Estimates Be Obtained?	11
1.4.2 Randomized and Bayesian Approaches	18
1.5 Randomization for Reducing Computational Complexity	18
1.6 Quantum Computing	19
2 Historical Overview	23
2.1 Game Theory	23
2.2 Monte Carlo Method, Random Search	24
2.2.1 Random Search, Simulating Annealing, Genetic Algorithms	25
2.2.2 Probabilistic Methods in a Control Syntheses	27
2.3 Estimation and Filtering under Arbitrary External Noises	28
2.3.1 Randomized Stochastic Approximation	28

VIII Contents

2.3.2	Linear Regression and Filtering	34
2.3.3	Compressive Sensing	36
2.3.4	Randomized Control Strategies	38
2.4	Data Mining and Machine Learning	40
2.4.1	Clustering	41
2.4.2	Cluster Validation	42

Part II: Randomization in Estimation, Identification and Filtering Problems

3	Randomized Stochastic Approximation	51
3.1	Mean-Risk Optimization	51
3.2	Exciting Testing Perturbation as Randomization: Estimation Algorithms	53
3.3	Convergence of Estimates	54
3.4	Fastest Rate of Convergence	56
3.5	Tracking	58
3.6	Algorithm Implementation and Quantum Computing	61
3.7	Applications	63
3.7.1	Optimization of a Server Processing Queue	63
3.7.2	Load Balancing	69
3.7.3	UAV Soaring	71
4	Linear Models	75
4.1	Linear Regression and Filtering under Arbitrary External Noise	75
4.1.1	Linear Regression	75
4.1.2	Application in Photoemission Experiments	79
4.1.3	Moving Average	83
4.1.4	Filtering	84
4.2	Random Projections and Sparsity	88
4.2.1	Compressed (Spars) Signals	88
4.2.2	Transforming Coding	89
4.2.3	Compressive Sensing	91
4.2.4	Universal Measurement Matrix: Random Projections	92
4.2.5	Reconstruction Algorithms through ℓ_1 - Optimization and Others	96
4.2.6	Some Applications	101
5	Randomized Control Strategies	107
5.1	Preliminary Example	108
5.2	Problem Statement	112
5.3	Control Actions with Randomized Test Signals	113

5.4	Assumptions and Model Reparameterization	114
5.5	Stochastic Approximation Algorithm	116
5.6	Procedure for Constructing Confidence Regions	118
5.7	Combined Procedure	123
5.8	Randomized Control for Small UAV under Unknown Arbitrary Wind Disturbances	123

Part III: Data Mining

6	Clustering	131
6.1	Partition into k Clusters	131
6.2	k -Means Clustering	133
6.2.1	Randomized Iterative k -Means	134
6.2.2	Example	136
6.3	Clustering and Compressive Sensing	137
6.4	Gaussian Mixtures and Clustering	138
6.5	Information Methods	143
6.5.1	Mixture Clustering Model — An Information Standpoint	143
6.5.2	Information Bottleneck	144
6.6	Spectral Clustering	145
6.6.1	The Ng-Jordan-Weiss (NJW) Algorithm	148
6.6.2	σ -Learning and Model Selection	151
6.6.3	Numerical Experiments	152
6.6.4	Application to Writing Style Determination	156
7	Cluster Validation	163
7.1	Stability Criteria	163
7.1.1	External Indexes	163
7.1.2	The Cleft Algorithm	165
7.1.3	The General Stability Approach	167
7.2	Geometrical Criteria	168
7.3	Information Criteria	170
7.4	Probability Metric Method	172
7.4.1	Compound Metrics	174
7.4.2	Simple Metrics	175
7.4.3	Information Metrics	177
7.5	Simulation	178
7.5.1	First Simulation Method	178
7.5.2	Second Simulation Method	180
7.5.3	Third Simulation Method	180
7.6	Application of Kernel-Based Distances	181
7.6.1	Real Positive and Negative Definite Kernels	182
7.6.2	Two Sample Test Kernel Distances	185

7.6.3	Parzen Window-Based Distances	187
7.7	The Two-Sample Energy Test	189
7.7.1	Method Description	190
7.7.2	Numerical Experiments	191
7.8	Application of Minimal Spanning Tree Methodology	193
7.8.1	Friedman-Rafsky's <i>MST</i> Two-Sample Test	193
7.8.2	First Approach	194
7.8.3	Numerical Experiments	196
7.8.4	Second Approach	200
7.8.5	Experimental Results	204
7.9	Binomial Model and the $K - NN$ Approach	206
7.9.1	Methodology	209
7.9.2	Numerical Experiments	211
7.9.3	Comparison with Other Methods	215
7.10	Hotelling Metrics	216
7.10.1	Approach Description	216
7.10.2	Experimental Results	217
7.11	Cluster Validation as an Optimization Problem	222
7.11.1	Fast Randomized Algorithm for Finding True Number of Clusters	224
7.11.2	Numerical Experiments	226
A	Appendix— Experimental Data Sets	229
A.1	Synthetic Data	229
A.1.1	G5: Five-Cluster Simulated Gaussian Dataset	229
A.2	Real-World Data	230
A.2.1	Three-Text Collections	230
A.2.2	Three-Text Collection—Second Version	230
A.2.3	Iris Flower Dataset	230
A.2.4	The Wine Recognition Dataset	231
A.2.5	The Glass Dataset	231
A.2.6	The Flea-Beetles Dataset	231
	References	233
	Index	247